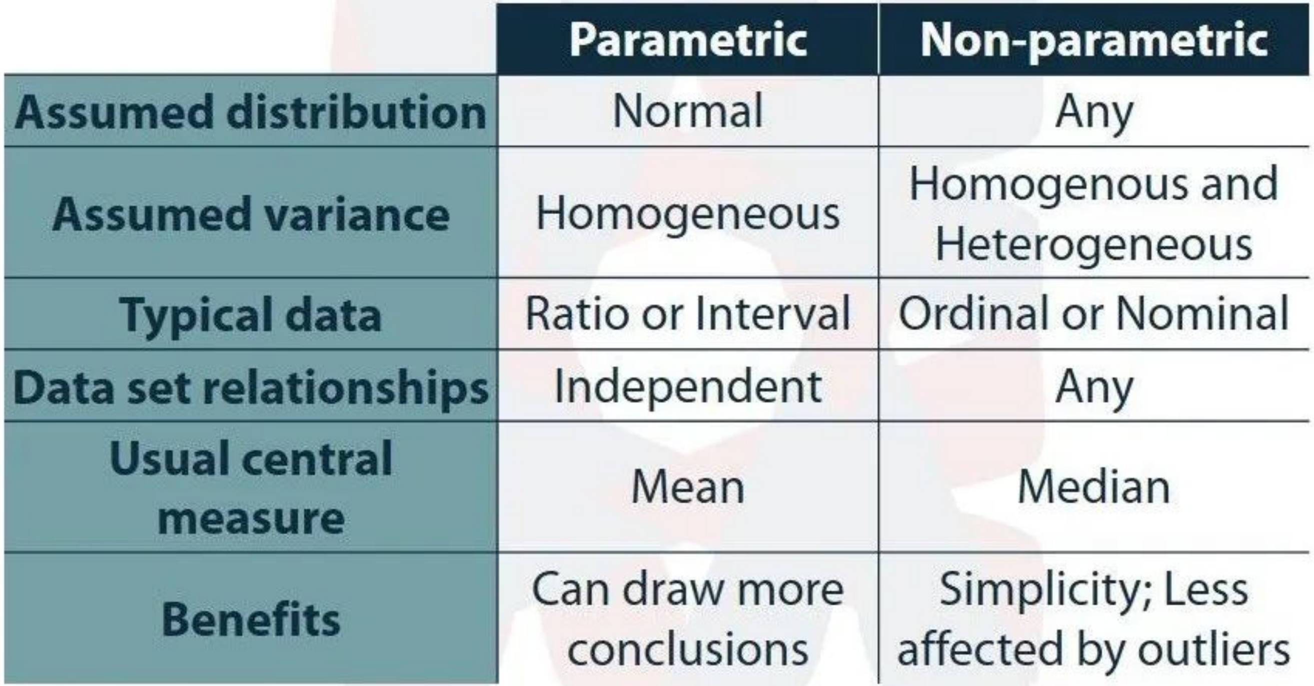
Chi-Square Analysis

Non-Parametric Test

Parametric vs Non-Parametric



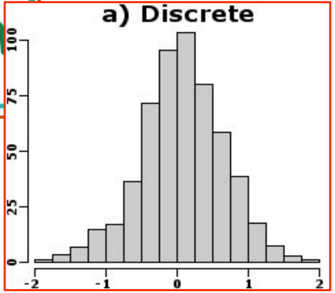
	Parametric	Non-parametric
Assumed distribution	Normal	Any
Assumed variance	Homogeneous	Homogenous and Heterogeneous
Typical data	Ratio or Interval	Ordinal or Nominal
Data set relationships	Independent	Any
Usual central measure	Mean	Median
Benefits	Can draw more conclusions	Simplicity; Less affected by outliers

Types of Variables

Quantitative

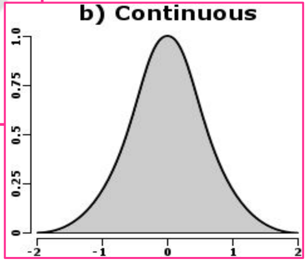
Discrete

# of children, # of patients



Continuous

height, weight, age

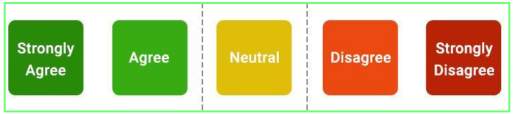


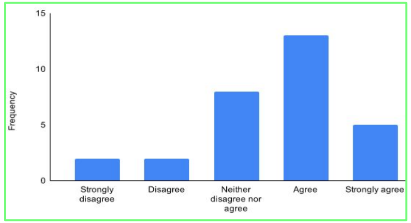
Types of Variables

Categorical

Ordinal

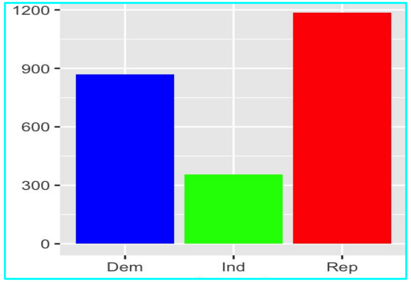
grades, Likert scale, ADL





Nominal

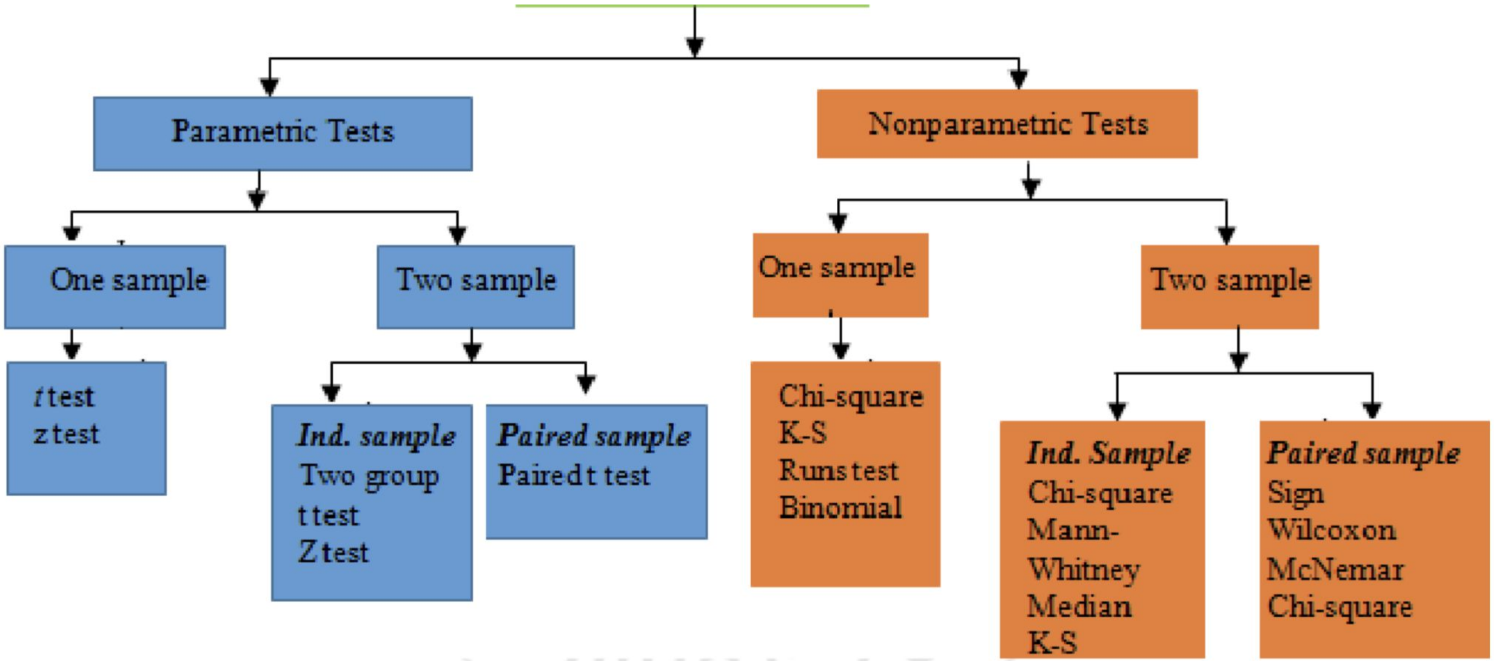
gender, race, make



Binary

yes/no; heads, tails

Parametric vs Non-Parametric



- Parametric Tests
  - One sample
    - *t* test
    - z test
  - Two sample
    - ***Ind. sample***
      - Two group
      - t test
      - Z test
    - ***Paired sample***
      - Paired t test
- Nonparametric Tests
  - One sample
    - Chi-square
    - K-S
    - Runs test
    - Binomial
  - Two sample
    - ***Ind. Sample***
      - Chi-square
      - Mann-Whitney
      - Median
      - K-S
    - ***Paired sample***
      - Sign
      - Wilcoxon
      - McNemar
      - Chi-square

Figure II: Classification of a one and two sample tests

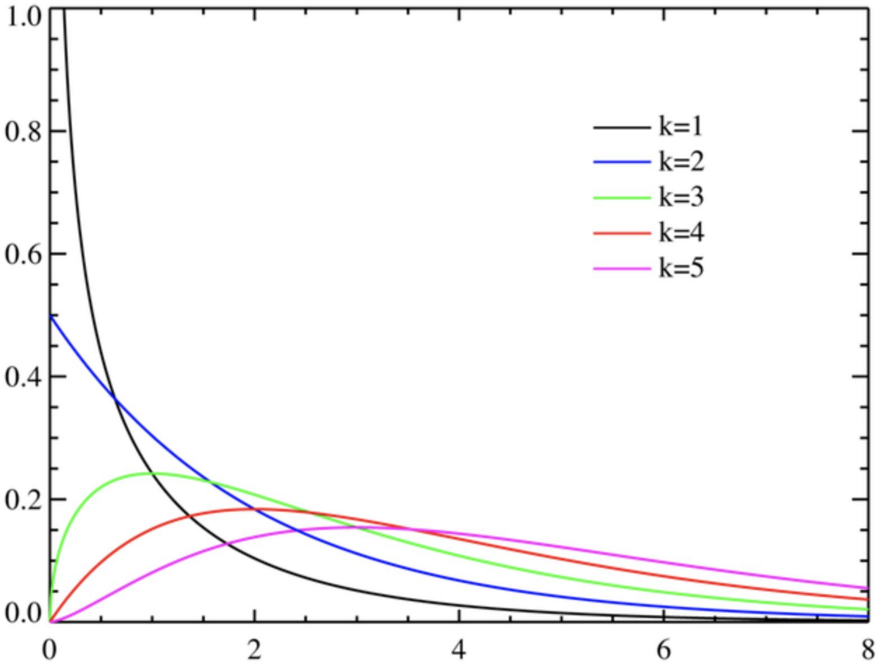
Chi Square (χ^2) Analysis

- Test the relationship between **categorical** variables
  - using distribution of numerical values
  - count, frequency, proportion
- Does the observed proportion match the expected proportion
  - in relation to the categories of a specific variable

$$\boxed{\frac{\text{Observed difference} - \text{Expectation if } H_0 \text{ is true}}{\text{Average Variance}}}$$

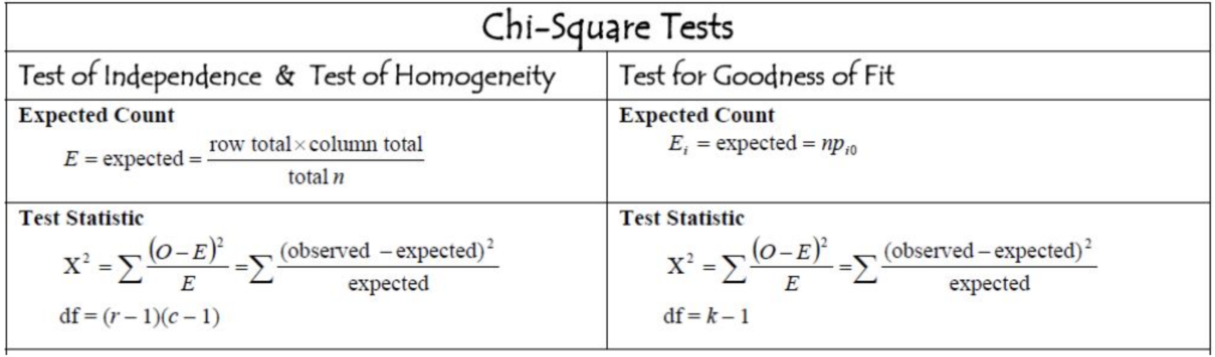
$$\chi^2 = \sum \frac{(Observed - Expected)^2}{Expected} = \sum \frac{(O-E)^2}{E}, \; df = k - 1$$

Chi Square (χ^2) Analysis



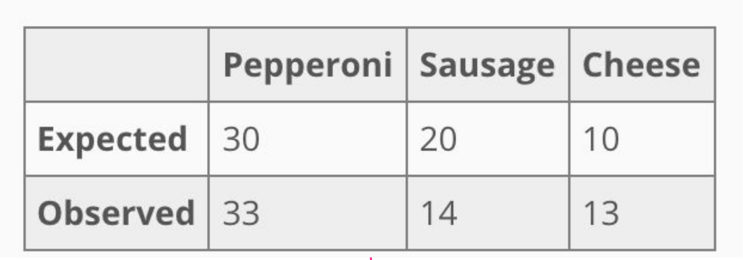
Two Types of Chi Square Analysis

1. **Test of Goodness -of-fit**
   - is observed frequency is different from the expected
   - e.g. *the occurrence of heads and tails while flipping a coin*
2. **Test of Independence**
   - is observed frequency is different from the expected
   - test if two variables are associated or independent
   - e.g. association between smoking and lung cancer



Chi-Square Tests	
Test of Independence & Test of Homogeneity	Test for Goodness of Fit
Expected Count $E = \text{expected} = \frac{\text{row total} \times \text{column total}}{\text{total } n}$	**Expected Count** $E_i = \text{expected} = np_{i0}$
Test Statistic $X^2 = \sum \frac{(O-E)^2}{E} = \sum \frac{(\text{observed} - \text{expected})^2}{\text{expected}}$ $\text{df} = (r-1)(c-1)$	**Test Statistic** $X^2 = \sum \frac{(O-E)^2}{E} = \sum \frac{(\text{observed} - \text{expected})^2}{\text{expected}}$ $\text{df} = k-1$

Chi Square (χ^2) Analysis



	Pepperoni	Sausage	Cheese
Expected	30	20	10
Observed	33	14	13

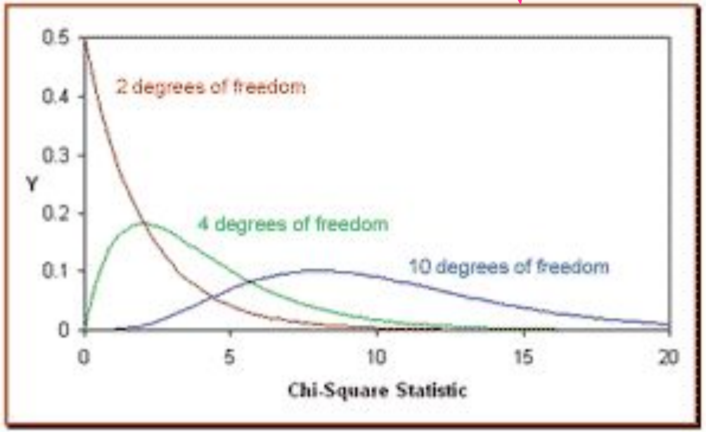
$$\chi^2 = \sum \frac{(o-e)^2}{e}$$

$$Pepperoni: \quad \frac{(33-30)^2}{30} = 0.3$$

$$Sausage: \quad \frac{(14-20)^2}{20} = 1.8$$

$$Cheese: \quad \frac{(13-10)^2}{10} = 0.9$$

chi square is 0.3 + 1.8 + 0.9 = 3



Chi Square (χ^2) Analysis

Income Level

Tax Bill	Low	Medium	High	
For	213 (209.9)	203 (187.2)	182 (200.9)	598
Against	138 (141.1)	110 (125.8)	154 (135.1)	402
	351	313	336	1000

H_0: voter's opinion concerning the new tax bill and their income level is independent

H_1: voter's opinion concerning the new tax bill and their income level is not independent

Test Statistic: $\chi^2 = \sum \frac{(Observed - Expected)^2}{Expected} = \sum \frac{(O-E)^2}{E}$, where $df = (2-1)(3-1) = 2$

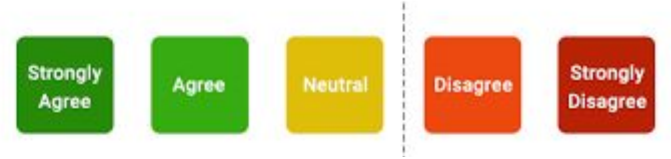
$$\chi^2 = \frac{(213-209.9)^2}{209.9} + \frac{(203-187.2)^2}{187.2} + \frac{(182-200.9)^2}{200.9} + \frac{(138-141.1)^2}{141.1} + \frac{(110-125.8)^2}{125.8} + \frac{(154-135.1)^2}{135.1} = 7.85$$

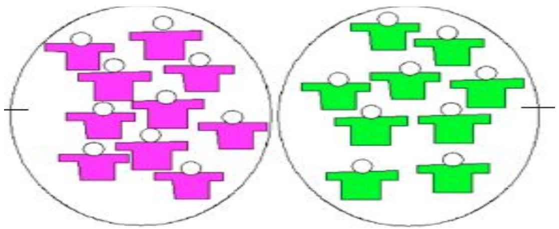
Critical Value: $\chi^2_a = \chi^2_{(0.05,2)} = 5.99$

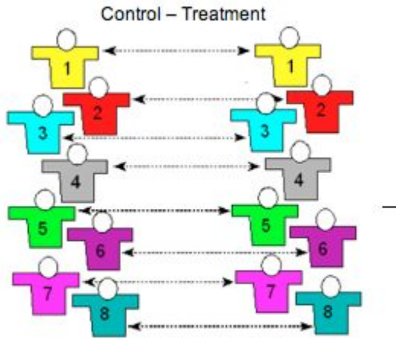
Sine $\chi^2 = 7.85 > \chi^2_a = 5.99$ we reject the null hypothesis and conclude that voter opinion concerning the new tax reform bill and income level are not independent.

Appendix

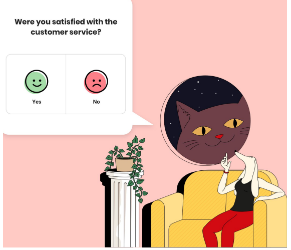
Non-Parametric Tests







Mann-U

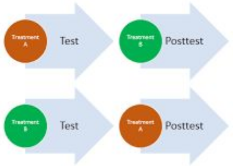

Were you satisfied with the customer service?
Yes
No

Kruskal-Wallis

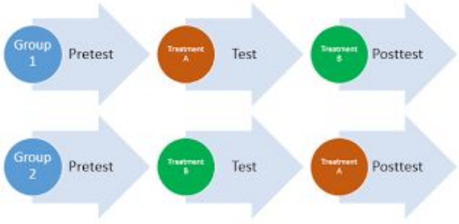
1. I feel like I belong in my company.

○ Strongly agree
○ Agree
○ Neither agree nor disagree
○ Disagree
○ Strongly disagree

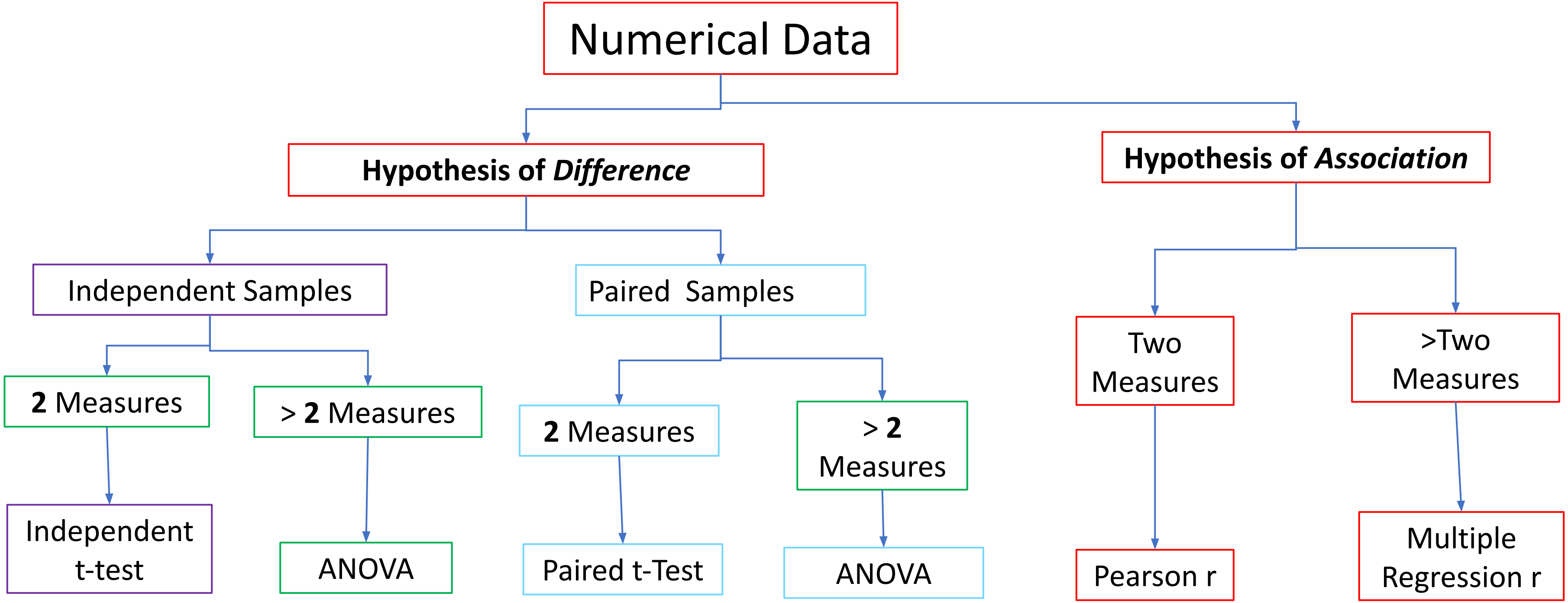
Wilcoxon



Friedman



Numerical Data



Numerical Data

Hypothesis of *Difference*

Hypothesis of *Association*

Independent Samples

Paired Samples

2 Measures

> **2** Measures

2 Measures

> **2** Measures

Two Measures

>Two Measures

Independent t-test

ANOVA

Paired t-Test

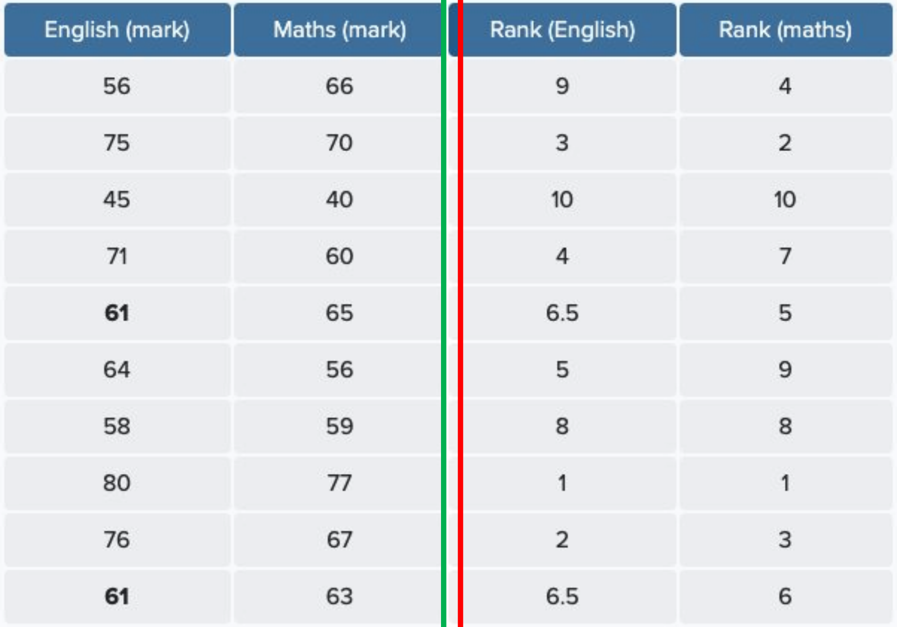
ANOVA

Pearson r

Multiple Regression r

Pearson vs Spearman Correlation

Pearson **Spearman-Rank**



English (mark)	Maths (mark)	Rank (English)	Rank (maths)
56	66	9	4
75	70	3	2
45	40	10	10
71	60	4	7
61	65	6.5	5
64	56	5	9
58	59	8	8
80	77	1	1
76	67	2	3
61	63	6.5	6

Spearman-Rank

$$\rho = 1 - \frac{6\sum d_i^2}{n(n^2-1)}$$

Pearson

$$\rho = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2}}$$